JOURNAL OF VIROLOGY, July 2010, p. 6497–6504 Vol. 84, No. 13
0022-538X/10/$12.00 doi:10.1128/JVI.01603-09

Genome-Scale Phylogenetic Analyses of Chikungunya Virus Reveal Independent Emergences of Recent Epidemics and Various Evolutionary Rates▽‡

Sara M. Volk,[1]†§ Rubing Chen,[1]† Konstantin A. Tsetsarkin,[1] A. Paige Adams,[1] Tzintzuni I. Garcia,[2]¶ Amadou A. Sall,[3] Farooq Nasar,[1] Amy J. Schuh,[1] Edward C. Holmes,[4] Stephen Higgs,[1] Payal D. Maharaj,[5]‖ Aaron C. Brault,[5]‖ and Scott C. Weaver[1]*

Center for Biodefense and Emerging Infectious Diseases and Department of Pathology,[1] and Department of Biochemistry and Molecular Biology,[2] University of Texas Medical Branch, Galveston, Texas; Institut Pasteur de Dakar, Dakar, Senegal[3]; Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, Pennsylvania[4]; and Center for Vector-Borne Diseases and Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, California[5]

Received 2 August 2009/Accepted 29 March 2010

***Chikungunya virus* (CHIKV), a mosquito-borne alphavirus, has traditionally circulated in Africa and Asia, causing human febrile illness accompanied by severe, chronic joint pain. In Africa, epidemic emergence of CHIKV involves the transition from an enzootic, sylvatic cycle involving arboreal mosquito vectors and nonhuman primates, into an urban cycle where peridomestic mosquitoes transmit among humans. In Asia, however, CHIKV appears to circulate only in the endemic, urban cycle. Recently, CHIKV emerged into the Indian Ocean and the Indian subcontinent to cause major epidemics. To examine patterns of CHIKV evolution and the origins of these outbreaks, as well as to examine whether evolutionary rates that vary between enzootic and epidemic transmission, we sequenced the genomes of 40 CHIKV strains and performed a phylogenetic analysis representing the most comprehensive study of its kind to date. We inferred that extant CHIKV strains evolved from an ancestor that existed within the last 500 years and that some geographic overlap exists between two main enzootic lineages previously thought to be geographically separated within Africa. We estimated that CHIKV was introduced from Africa into Asia 70 to 90 years ago. The recent Indian Ocean and Indian subcontinent epidemics appear to have emerged independently from the mainland of East Africa. This finding underscores the importance of surveillance to rapidly detect and control African outbreaks before exportation can occur. Significantly higher rates of nucleotide substitution appear to occur during urban than during enzootic transmission. These results suggest fundamental differences in transmission modes and/or dynamics in these two transmission cycles.**

Chikungunya virus (CHIKV; *Togaviridae*: *Alphavirus*) is an arbovirus (arthropod-borne virus) vectored by *Aedes* mosquitoes to humans in tropical and subtropical regions of Africa and Asia (Fig. 1; reviewed in references 26 and 46). CHIKV has a single-stranded, positive-sense RNA genome of ~12 kb and causes chikungunya fever (CHIK), a febrile illness associated with severe arthralgia and rash (2, 15, 31, 35); the name is derived from a Bantu language word describing the severe arthritic signs (32), which can persist for years. Thus, CHIK has enormous economic costs in addition to its public health impact (9). Because the signs and symptoms of CHIK overlap with those of dengue and because CHIKV is transmitted sympatrically in urban areas by the same mosquito vectors, it is grossly underreported in the absence of laboratory diagnostics (10, 37).

CHIKV was first isolated during a 1953 outbreak in present-day Tanzania by Ross (48, 49). Since then, outbreaks have been documented in Africa and Asia, including the Indian subcontinent (Fig. 1) (1, 4). In 2005, CHIKV emerged from East Africa to cause an explosive urban epidemic in popular tourist island destinations in the Indian Ocean (Fig. 1; reviewed in reference 31). In late 2005, CHIKV spread into the Indian subcontinent, where millions of people have been affected (5). However, the geographic source of spread into India, from the mainland of Africa or from the Indian Ocean Islands, has not been delineated. India had seen large epidemics of CHIK in the past (reviewed in reference 30), but CHIKV apparently disappeared during the 1970s (5). Since 2006, CHIKV has been imported into Europe and the western hemisphere (including the United States) via many viremic travelers, and an epidemic was initiated in Italy by a traveler from India (4, 11, 47). The dramatic spread since 1980 of dengue viruses (DENV) throughout tropical America, via the same vectors, portends the severity of the public health problem if CHIKV becomes established in the western hemisphere.

* Corresponding author. Mailing address: University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-0609. Phone: (409) 747-0758. Fax: (409) 747-2429. E-mail: sweaver@utmb.edu.

† S.M.V. and R.C. contributed equally to this study.

‡ Supplemental material for this article may be found at http://jvi.asm.org/.

§ Present address: Department of Biology, Our Lady of the Lake University, San Antonio, TX 78207.

¶ Present address: *Xiphophorus* Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, TX 78666.

‖ Present address: Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521.

▽ Published ahead of print on 21 April 2010.

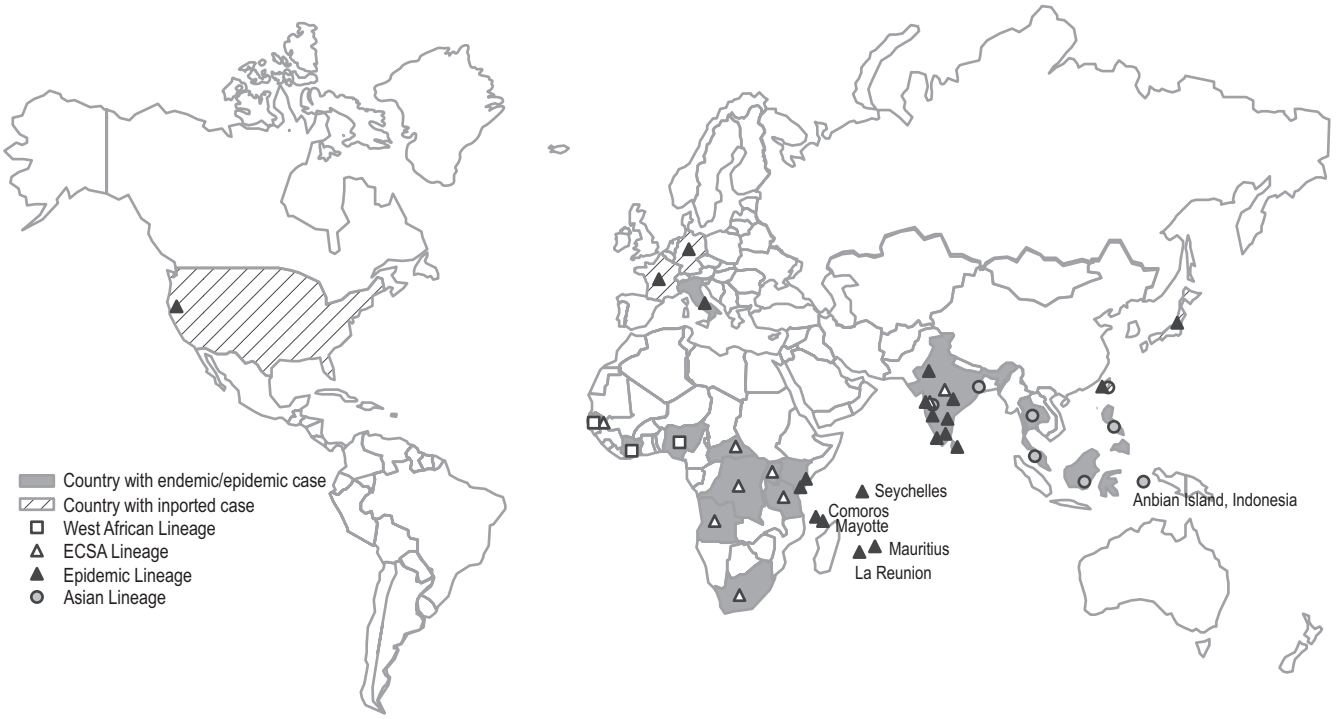

FIG. 1. Distribution of the CHIKV strains used in this study. The map, based on a world map template from http://www.presentationmagazine.com, was edited with permission.

The first phylogenetic analysis of CHIKV (45) identified three geographically associated genotypes: the West African (WAf), East/Central/South African (ECSA), and Asian genotypes. More recent analyses indicate that the recent Indian Ocean and Indian strains form a monophyletic group within the ECSA lineage (5, 12, 14, 27, 40, 51, 52). However, most CHIKV phylogenetic studies (1, 14, 28, 29, 38, 40, 41, 47, 52) have utilized only partial sequences from the envelope glycoprotein E1 gene, preventing a robust assessment of some of the relationships among strains and of their evolutionary dynamics.

The CHIKV strains represented in different geographic lineages apparently circulate in different ecological cycles. In Asia, CHIKV appears to circulate primarily in an urban transmission cycle involving the peridomestic mosquitoes *Aedes aegypti* and *A. albopictus*, as well as humans (25, 45). Asian epidemics typically infect thousands-to-millions of people over the course of several years (46). In contrast, African CHIKV circulates primarily in a sylvatic/enzootic cycle, transmitted by arboreal primatophilic *Aedes* mosquitoes (e.g., *A. furcifer* and *A. africanus*) and probably relies on nonhuman primates as reservoir hosts (reviewed in reference 16). Epidemics in rural Africa usually occur on a much smaller scale than in Asia, likely a result of the lower human population densities, and possibly more stable herd immunity. Although the assignments of "urban" and "sylvatic/enzootic" are based on the most common mode of transmission, CHIKV strains of African origin are capable of urban transmission by *A. aegypti* and *A. albopictus*, as evidenced by outbreaks in the Democratic Republic of the Congo (41), Nigeria (36), Kenya (27), and Gabon (42). The ecological differences between the sylvatic/enzootic (henceforth called enzootic) and urban/endemic/epidemic transmission cycles (henceforth called epidemic) such as seasonality of vector larval habitats, vertebrate host abundance and herd immunity, and vector host preferences, prompted us to hypothesize that the evolutionary dynamics of CHIKV may differ between the two transmission cycles. To test this hypothesis, to provide more robust estimates of the evolutionary relationships among the CHIKV strains including the sources of the recent epidemics, and to elucidate the temporal and spatial history of CHIKV evolution, we performed an extensive, genome-scale phylogenetic analysis, utilizing complete open reading frame (ORF) sequences of a large collection of 80 isolates with broad temporal, spatial, and host coverage.

MATERIALS AND METHODS

Virus samples. CHIKV strains listed in Table S1 in the supplemental material were either obtained from the World Reference Center for Emerging Viruses and Arboviruses at the University of Texas Medical Branch, Galveston, TX, or from the Institut Pasteur de Dakar, Senegal. Viruses were passaged in C6/36 *A. albopictus* cells, concentrated with polyethylene glycol (7% [wt/vol]) and NaCl (2.3% [wt/vol]), and RNA was extracted by using either TRIzol (Invitrogen, Carlsbad, CA) or the QIAamp viral RNA minikit (Qiagen, Valencia, CA) according to the manufacturers' protocols.

RT-PCR and sequencing. Eight or nine overlapping PCR amplicons were generated from viral RNA using the Titan One Tube reverse transcription-PCR system (Roche, Mannheim, Germany) according to the manufacturer's protocol. The amplicons were subsequently gel purified and sequenced by using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA). Sequencing was performed in an ABI Prism model 3100 genetic analyzer (Applied Biosystems, Foster City, CA), and the sequences were edited and assembled in Sequencher v4.9 (Gene Codes, Ann Arbor, MI), and deposited in the GenBank database (see Table S1 in the supplemental material). Primer sequences and specific PCR and sequencing protocols are available from the authors.

Phylogenetic analyses. Newly generated CHIKV genomic sequences, as well as those available from the GenBank library (excluding vaccine and cloning vector strains), along with that of o'nyong-nyong virus (ONNV; strains Gulu, Igbo Ora, and SG650, used as an outgroup) were aligned by using MUSCLE (20) and manually adjusted in Se-Al (available at http://tree.bio.ed.ac.uk/software/seal/) according to amino acid sequence alignments to preserve codon homology. Due to the ambiguous alignments in the 3′ untranslated region (3′UTRs) and rapid evolution in the other UTR, only open reading frames (ORFs) were adopted in further analysis. This led to an alignment of 11,319 nucleotides, containing 80 CHIKV strains and 3 ONNV isolates. To search for any potential recombination event in CHIKV that could affect the phylogenetic structure, genomic sequences were screened exhaustively and triplet-by-triplet in RDP version 3.41 (34) and using a suite of different recombination detection methods using RDP (33), Chimaera (43), MaxChi (57), 3Seq (8), and GENECONV (39). Recombination with significantly positive hits was not found in our data set, including in the high passage isolates.

Phylogenetic trees were inferred by using both the maximum-likelihood (ML) method in the PAUP* v4.0b package (58) and the Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) method in MrBayes v3.1.2 (24). MODELTEST (44) was used to select the best-fit model for the ML analyses. MrBayes analyses used the GTR+I+Γ_4 model; in addition, the nucleotide data were partitioned by three codon positions, and substitution parameters were allowed to vary across partitions. The analysis used three hot chains and one cold chain and ran for 10 million generations with 25% burnin; the "sump" command and Tracer v1.4.1 (http://tree.bio.ed.ac.uk/software/tracer/) were used to ensure samples were taken after likelihoods had stabilized.

To examine the possible advantage of using complete genome sequences in the phylogenetic study, we also inferred an ML tree based on the E1 gene and compared it to the ML phylogeny of complete ORF. To determine whether these trees differ significantly in topology, we used the Shimodaira-Hasegawa (SH) test implemented in PAUP.

Rates of nucleotide substitution and ttMRCA. We used the Bayesian Markov Chain Monte Carlo (MCMC) method available in BEAST v1.5.3 (18) to estimate evolutionary rates and times to the most recent common ancestors (tMRCA) for CHIKV overall and for each of the individual clades, namely, WAf, ECSA, Asian, and the recent epidemic group. To avoid artifacts due to laboratory adaptation, high-passage strains (Ross and S27) were excluded from this analysis, as well as those without clear sample year information (see Table S1 in the supplemental material). The strains Angola/M2022/1962, India/MH4/2000, and India/ALSA-1/1986 were also omitted because of the potential that they were the result of contamination or high passage, as suggested by either a suspicious terminal branch length or phylogenetic position (see Fig. S1 and S2 in the supplemental material). However, strains with low passage histories (<10) were used in analysis because previous studies with alphaviruses have shown that these small numbers of passages introduce few mutations (13, 62) and should therefore have little or no effect on coalescent studies. All strains were dated according to the year and month (if known) of their collection; strains with only a known collection year were assigned the median-year value for July. Due to the short time scale of the recent Indian Ocean basin epidemic group, only sequences with a known sample month were included in evolutionary rate and tMRCA estimations. This led to a total data set of 80 genomic CHIKV sequences, including 21 Asian, 11 WAf, 9 ECSA, and 39 recent epidemic strains. BEAST analysis was performed for the combined data set and for individual clades, based on a relaxed molecular clock (uncorrelated lognormal) and the SRD06 nucleotide substitution model that has been suggested to be superior to other models when used with ORFs in RNA viral genomes (54). All analyses also used a Bayesian skyline coalescent tree prior (19), which imposes the fewest demographic assumptions, and because estimating demographic history was not an aim of the present study. In each case, MCMC chains were run for a sufficient time to achieve convergence (accessed using the Tracer program; http://tree.bio.ed.ac.uk/software/tracer). Statistical uncertainty in parameter estimates is reflected as the 95% highest probability density (HPD) values. The maximum clade credibility tree across all of the plausible trees generated by BEAST was then computed by using the TreeAnnotator program available in BEAST package, with the first 10% of trees removed as burn-in. To assess the reliability of our substitution rate and tMRCA estimates and to determine the extent of the temporal structure of the sequence data, we also performed a regression analysis of tree root-to-tip genetic distance against sampling dates using the program Path-O-Gen (http://tree.bio.ed.ac.uk/software/pathogen/) (17) for each lineage based on the corresponding phylogeny excised from the ML tree of 80 CHIKV strains.

Selection pressures. To investigate the nature of selection pressures that may act on CHIKV, we estimated the average numbers of nonsynonymous (*dN*) and synonymous (*dS*) nucleotide substitutions per site (*dN*/*dS* ratio) for each clade using the "one-ratio" model from the CODEML program available within the PAML package (63), using the subtrees of the total ML tree indicated above. In addition, the selection pressures on external and internal branches of each ML phylogeny were estimated separately using the "two-ratio" model available in CODEML. Similarly, the numbers and locations of positively and negatively selected sites were estimated by using the internal fixed effects likelihood method available in the HYPHY package (42a).

RESULTS

Genetic diversity of CHIKV. Forty CHIKV genomes were sequenced and aligned with those available in GenBank. Excluding the 5′ and 3′ 20 nucleotides (nt) that were not sequenced, the genome length varied among and within geographic lineages, with those in the ECSA lineage being shorter (11,557 to 11,789 nt) than the WAf (11,843 to 11,881 nt) and Asian (11,777 to 11,999 nt) strains. Nucleotide differences were found in all genes, and the most variable genome regions included the 5′ and 3′ UTRs, as well as the 26S junction region. The ORFs were highly conserved, with occasional indels observed in high-passage strains. The highly divergent UTRs made accurate alignments impossible, and the UTRs were therefore excluded from the phylogenetic analyses. We found poly(A) insertions in the 3′UTRs of two ECSA strains (Fig. 2) in addition to the Ross strain (5). The insertion in the Senegal bat strain, which is located in the ECSA group, is in a different location from the other two, suggesting their independent generation.

Origin and divergence of geographically related clades. Similar to previous findings (47), our phylogenetic trees (shown in Fig. 2 and see Fig. S1 and S2 in the supplemental material) all included three distinct CHIKV clades, namely, ECSA, West Africa, and Asian, with the recent Indian Ocean basin outbreak forming a monophyletic lineage descendant from the ECSA clade. The divergence of each distinct lineage reflected, to some extent, the path of global transmission and occasional outbreaks. According to our estimate, the currently circulating CHIKV strains have an ancestor that existed within the last 500 years, with 95% HPD values of the most recent common ancestor (MRCA) ranging from 169 to 436 years ago. Interestingly, despite their close geographic distance, the two African lineages did not cluster together, indicating limited genetic exchange between the two lineages in Africa. The only exception was a 1963 bat isolate from Senegal, which grouped in the ECSA clade. This finding is the first to suggest that the main West African and ECSA lineages may overlap spatially in the enzootic cycle, at least on occasion.

The divergence of the ECSA and Asian lineages occurred within the last 150 years (95% HPD: 1879 to 1927 AD), with the Asia group splitting into two clades: an Indian lineage, which likely went extinct (5, 30), and a Southeast Asian lineage, which is probably still circulating. The Southeast Asian lineage suggests a remarkable spatial and temporal pattern, spreading from Thailand to Indonesia, and then to the Philippines, and more recently to Malaysia (Fig. 2). This temporal branching pattern was similar to patterns observed in some DENV phylogenies (23) and in other alphaviruses such as eastern equine encephalitis virus in North America (6).

Similarly, the recent Indian Ocean basin outbreak that began in 2004 apparently originated from the ECSA group as

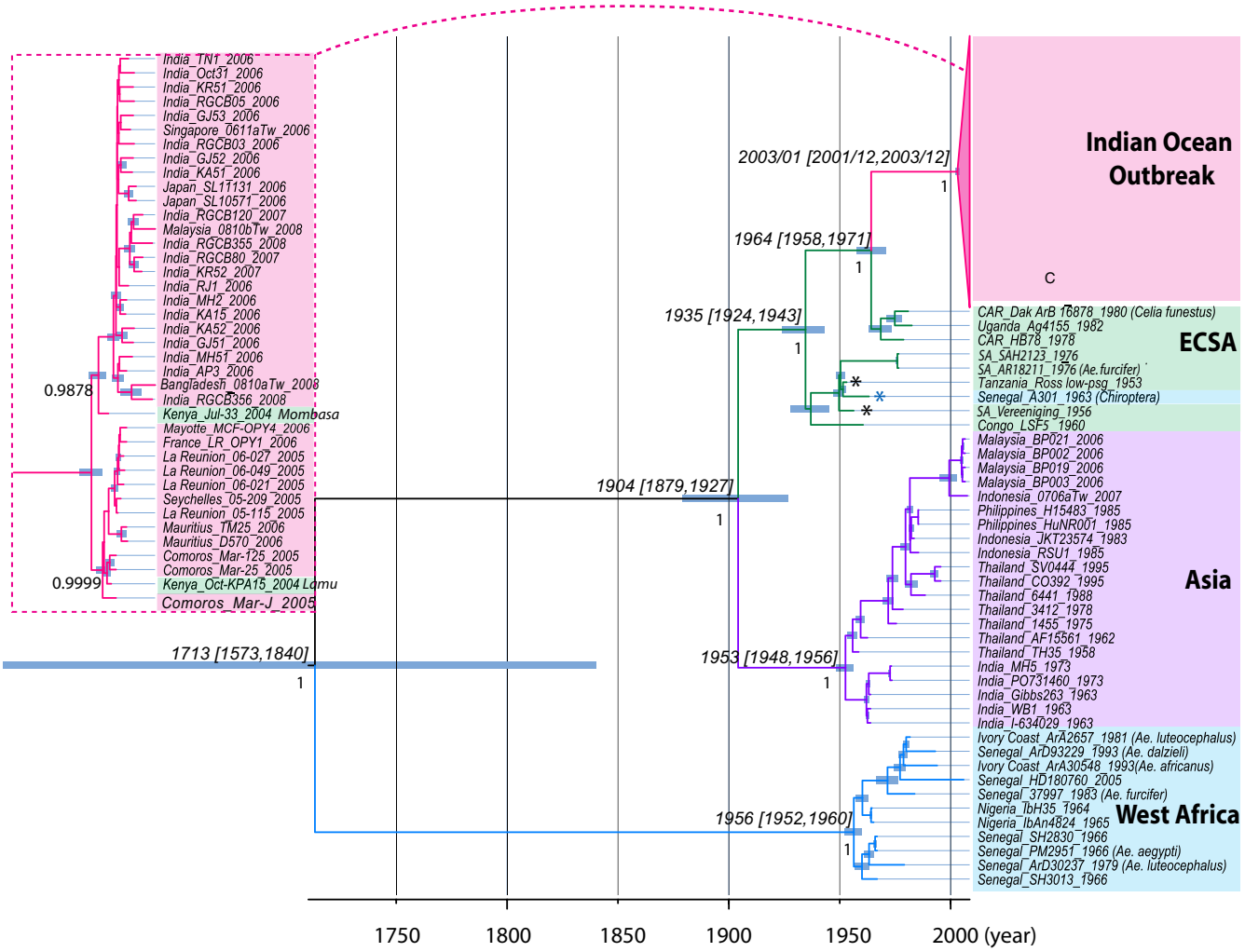

FIG. 2. Maximum clade credibility (MCC) tree of 80 CHIKV strains. The four major lineages are highlighted with different branch colors, with the sample origin highlighted by the corresponding color. The estimated 95% HPD values for most recent common ancestors are labeled beside the node and are also indicated by the thick blue horizontal node bars. The numbers adjacent to nodes indicate Bayesian posterior probability values. For clarity, the 2004 to 2009 epidemic clade is shown enlarged in the inset at upper left. Sequences with internal poly(A) insertion in the 3′UTR are marked with an asterisk (*). Strains are labeled as follows: location_strain name_date (year) of collection.

shown in Fig. 2, with the MRCA dating back to about 2002 (95% HPD: December 2001 to December 2003). Two distinct lineages were identified for the Indian Ocean and Indian subcontinent strains (as well as strains exported from these regions). The Indian subcontinent clade was rooted by a mainland Kenya 2004 strain, while the Indian Ocean clade also contains a 2004 Kenyan strain (Lamu Island), indicating independent emergences from East Africa into the Indian Ocean and the Indian subcontinent.

The ECSA clade has been sampled in a broad geographic range since the first identification of CHIKV in 1953, with the MRCA of available isolates dating to from 1924 to 1943 (95% HPDs). In contrast to the clear pattern of geographic spread exhibited by the Asian lineage and the recent Indian Ocean epidemic clades, there was no clear geographic pattern of ECSA lineage spread. Due to the limited sample size, it is not clear in what mode and to what extent this lineage has been circulating in eastern, central, and southern Africa.

Interestingly, a mosquito isolate, MH4 from 2000, which was collected in India, fell within the ECSA clade (see Fig. S1 and S2 in the supplemental material), suggesting a CHIKV introduction from Africa into India at least 5 years before the 2005 epidemic. However, the high (99.4%) nucleotide identity between Uganda/Ag4155/1982 and India/MH4/2000 was inconsistent with their nearly 20-year difference in collection dates, given our estimates of CHIKV nucleotide substitution rates (Table 1). Given the overall strong temporal pattern shown in the phylogenetic trees, we suspect that India/MH4/2000 might be the result of contamination.

Evolutionary patterns among CHIKV lineages. The evolutionary rates estimated by using the Bayesian MCMC method for the complete CHIKV data set and individual lineages are summarized in Table 1. The overall nucleotide substitution rate was estimated as 4.33×10^{-4} nucleotide substitutions per site per year (subs/nt/year). However, the rates estimated for each lineage exhibited considerable variation, with those for

TABLE 1. Rates of nucleotide substitution of chikungunya virus

Parameter	CHIK80	Asian21	ECSA9	WAf11	Epidemic39
Mean rate	4.33E-04	4.16E-04	2.30E-04	2.39E-04	8.41E-04
95% HPD lower	3.15E-04	3.26E-04	1.37E-04	1.98E-04	5.78E-04
95% HPD upper	5.62E-04	5.02E-04	3.24E-04	2.84E-04	1.09E-03
Coefficient of variation	0.44	0.63	0.11	0.08	0.44

the epidemic lineages significantly higher than those estimated for the enzootic lineages. In particular, the Asian lineage exhibited a significantly higher substitution rate (i.e., nonoverlapping HPD values) (4.16×10^{-4} subs/nt/yr; 95% HPD: 3.26 to 5.02×10^{-4} subs/nt/year) than the WAf (2.39×10^{-4} subs/nt/year; 95% HPD: 1.98 to 2.84×10^{-4} subs/nt/year) and ECSA lineages (2.30×10^{-4} subs/nt/year; 95% HPD: 1.37 to 3.24×10^{-4} subs/nt/year). The Indian Ocean epidemic lineage yielded an even higher rate estimate (8.46×10^{-4} subs/nt/year; 95% HPD: 5.81×10^{-4} to 1.09×10^{-3} subs/nt/year). This widespread rate variation was confirmed by using root-to-tip linear regression, with all lineages showing strong clocklike behavior (as reflected in the correlation coefficient values; see Fig. S3 in the supplemental material).

It is possible, however, that the intensive sampling of sequences within a short time span from the recent epidemic may have included many transient deleterious mutations (i.e., sequence polymorphisms) that would later be purged by purifying selection and therefore not persist over longer time periods. This, in turn, would artificially increase the substitution rate for the Indian Ocean lineage. To test this hypothesis, we measured selection pressures among the four major CHIKV lineages. This analysis revealed a significantly higher overall *dN*/*dS* value in the Indian Ocean epidemic lineage (0.285) compared to the others (0.066 to 0.125), which is compatible with the presence of transient, mildly deleterious mutations in the former group. A similar hypothesis was proposed to explain the higher *dN*/*dS* values for swine-origin influenza A (H1N1) virus during the recent epidemic, compared to related swine influenza virus sequences (56). In further support of this hypothesis, the Indian Ocean epidemic sequences also exhibited an elevated number of nonsynonymous changes on external branches of the tree, reflected in a ratio of internal/external *dN*/*dS* values of 0.97, compared to the lower values observed in other groups (0.53 to 0.63; Table 2). In theory, transient deleterious mutations are more likely to fall in the external branches of trees (because they are short-lived), leading to a higher *dN*/*dS* value than seen for internal branches. Although *dN*/*dS* ratios have limited utility in identifying positively selected codons if unique, adaptive mutations occur, three positively selected codons were also observed in the epidemic lineage, comprising two codons in the capsid protein gene (codons 23 and 27) and one in the E1 envelope glycoprotein gene (codon 226); the latter plays a crucial role in CHIKV adaptation to *A. albopictus* (59, 61). However, these adaptive mutations can only partially explain the elevated *dN*/*dS* value in this lineage, since only 16 negatively selected codons were observed in the epidemic lineage, whereas many more (i.e., 64 to 96) were observed in other groups. Therefore, the elevated evolutionary rate in the Indian Ocean epidemic lineage is most likely due to the presence of transient deleterious mutations not seen in the other lineages. Finally, it was noteworthy that *dN*/*dS* was also elevated in the Asian group compared to the ECSA and WAf lineages (Table 2). This may indicate different selection pressures acting on the epidemic versus enzootic transmission cycles.

Comparison of complete genomic and partial E1 sequences in phylogenetic reconstructions. Importantly, the ML tree based on E1 gene sequences revealed a topology different from that based on the complete ORFs, especially for the ECSA lineage, which did not form a monophyletic group in the former data set (see Fig. S3 in the supplemental material). A significant difference between the E gene and complete ORF tree topologies was also apparent in the SH test ($P < 0.001$). Since the complete ORFs tree is clearly more accurate, possessing more variable and phylogenetic informative sites, this analysis shows that the E gene alone is inadequate to fully resolve the phylogenetic history of CHIKV.

DISCUSSION

CHIKV origins. CHIKV has likely been circulating in Africa and Asia for hundreds of years or even longer. A suspected CHIK epidemic was reported in 1779 (10), and focal epidemics have been documented occasionally throughout the second half of the 20th century. Recently, large-scale outbreaks successively swept through eastern Africa, the western Indian Ocean islands, India, and southeastern Asia and also reached Australia and Europe. These epidemics demonstrate the threat of this reemerging arbovirus and indicate the need to better understand its evolutionary history and patterns and, particularly, whether lineages differ in transmission cycles and ecological conditions, which could affect emergence potential.

Previous studies suggested a likely African origin of CHIKV (45), where the virus circulates in an enzootic cycle between forest-dwelling *Aedes* species mosquitoes and nonhuman primates. However, due to the deep divergence of the WAf and

TABLE 2. Selection pressures acting on each lineage of chikungunya virus

Data set	Mean *dN*/*dS* (95% CI)			I/E[a]	No. of negatively selected codons[b]	Positively selected codon position[b]
	Overall	Internal	External			
Asian21	0.125 (0.108–0.142)	0.095 (0.073–0.118)	0.152 (0.125–0.179)	0.63	70	NA
WAf11	0.066 (0.051–0.082)	0.040 (0.016–0.063)	0.075 (0.056–0.095)	0.53	64	NA
ECSA9	0.088 (0.060–0.115)	0.058 (0.049–0.068)	0.103 (0.080–0.126)	0.57	96	NA
Epidemic39	0.285 (0.232–0.338)	0.278 (0.056–0.497)	0.286 (0.300–0.343)	0.97	16	C-23; C-27; E1-226

[a] I/E/, ration of internal to external values.
[b] Significant at $\alpha = 0.05$.

ECSA lineages and the wide distribution in both East and West Africa of the closest relative, ONNV, it is unclear in which part of Africa CHIKV first evolved and, more intriguingly, what causes the nearly complete transmission barrier that separates evolution in the two geographic regions. This question deserves further study given that (i) there are no obvious geographical barriers or ecological differences and (ii) the two regions share the similar enzootic vector fauna and potential primate reservoirs and (iii) given the occasional but apparently nonpersistent mixing observed in the present study (the presence in Senegal of the ECSA clade, isolated from bats).

Epidemic origins. Interestingly, the Asian lineage dating to ca. 1952 exhibits similar patterns of spread to the recent Indian Ocean outbreak lineage, with successive epidemics detected along an eastward path. This similarity suggests that the current Asian lineage may have also originated from an intercontinental transmission. Indeed, the tMRCA of the old Asian lineage (95% HPD: 1948 to 1956) almost immediately predates the Thailand outbreak of 1958, the first confirmed CHIK epidemic in Asia. However, our current samples do not reveal exactly where the progenitor of the 1953 Asian outbreak occurred, although it was probably an ECSA strain that diverged from the currently circulating and sampled ECSA strains in the beginning of the last century (95% HPD: 1879 to 1927).

The repeated CHIKV outbreaks that have occurred in India and Southeast Asia are particularly noteworthy. In addition to epidemics corresponding to the Asian and Indian Ocean lineages, CHIKV outbreaks probably have occurred in Asia for more than 200 years (10). These outbreaks were characterized by rapid progression, affecting hundreds to millions of people, and were followed by long, silent interepidemic periods. Interestingly, CHIKV apparently did not persist in these areas, with the new outbreaks usually caused by introduction of imported strains, probably on sailing ships. The permanent Asian establishment of CHIKV during the 1950s may be attributed to the increased human urbanization and expansion of *A. aegypti* populations after World War II (21). The succession of CHIKV outbreaks in India and Southeast Asia is likely related to the decline of herd immunity in human populations, as shown by the long epidemic interval (e.g., 33 years since the last Indian circulation). In contrast, CHIK in Africa is endemic/enzootic, associated with limited, sporadic outbreaks (46), probably due to the lower densities of human populations and their relative isolation in many cases, as well as more stable herd immunity from periodic enzootic spillover.

Variation in evolutionary patterns. We hypothesized that the difference in vectors and amplification hosts between the epidemic and enzootic transmission cycles may influence CHIKV evolutionary rates, which was supported by our data. Although a previous analysis of evolutionary rates in DENV-2 virus, which also has ancestral enzootic and derived epidemic lineages, did not reveal faster evolution in the latter cycle, the number of enzootic DENV strains analyzed in this case was very small, with resultant broad 95% HPD values (60).

A variety of factors could contribute to the variation in evolutionary rate that we observed for CHIKV, involving differences in intrinsic factors, such as the rates of mutation and replication, or extrinsic factors, such as the strength of natural selection or population transmission rates. Because there is no evidence for differences in either mutation or replication rates among CHIKV lineages, we propose that the differences in evolutionary rate between the enzootic and epidemic lineages are more likely due to the different transmission rates (as discussed below), which affect the total amount of replication per unit time, and/or differences in selection pressure.

The replication of alphaviruses in mosquitoes depends initially on ambient temperature (reviewed in reference 22) and is eventually modulated by poorly understood mosquito factors, including RNA interference (50). The resulting RNA replication shutdown presumably leads to virtual genetic stasis despite the continued ability of infected vectors to transmit. Therefore, the transmission rate directly influences the amount of viral replication and consequently the numbers of mutations produced during the transmission cycle. The differing *dN*/*dS* values we estimated among epidemic versus enzootic CHIKV lineages, which may reflect differences in mildly deleterious mutations sampled intensively during the recent CHIKV epidemics, could partially explain the higher evolutionary rate of this 2004 to 2008 epidemic lineage. However, different transmission patterns between the epidemic and enzootic cycles may be critical in regulating evolutionary rates in the two cycles (as illustrated in Fig. 3). In the relatively stable epidemic cycle, CHIKV is transmitted among humans via abundant peridomestic mosquitoes, *A. aegypti* and *A. albopictus*, which colonize artificial and natural water containers in suburban and rural areas (53). In contrast, transmission in the enzootic cycle differs dramatically between wet and dry seasons, with fluctuations in vector densities associated with rainfall. During the dry season, due to the low vector population sizes, horizontal transmission of CHIKV is probably maintained at a low level, whereas transovarial transmission may play a role in CHIKV maintenance, as shown for other arboviruses (7). In addition, lower vertebrate host availability in the enzootic cycle, where primate populations are less dense than are humans in most urban habitats, may also lead to a lower transmission rate. Moreover, greater herd immunity of reservoir hosts in the enzootic cycle could also reduce transmission efficiency.

It is also possible that the lower evolutionary rates in enzootic lineages reflect stronger purifying selection, as suggested by the lower *dN*/*dS* values in the enzootic compared to the epidemic lineages (Table 2), which may in turn be related to the more diverse hosts and vectors used in the enzootic transmission cycle. Enzootic CHIKV probably circulates in several different nonhuman primates, and possibly in bats or other mammals, and is probably transmitted by several different *Aedes* mosquitoes in the *Celia* subgenus (shown in Fig. 2) (16). These diverse hosts and vectors may constrain the evolution and adaptation of enzootic CHIKV compared to the epidemic lineages that rely only on humans and 2 closely related *Aedes* (*Stegomyia*) vectors. Another possibility is that epidemic transmission involves more CHIKV population bottlenecks if humans generate lower viremia than nonhuman primates, leading to smaller numbers of virions establishing infection of the vector. However, if *A. aegypti* and/or *A. albopictus* are less susceptible than enzootic vectors, smaller numbers of virions may infect the midgut. Similar effects could occur if different amounts of virus are transmitted in the saliva of epidemic versus enzootic vectors. These bottlenecks could accelerate rates of sequence change by allowing the fixation of slightly

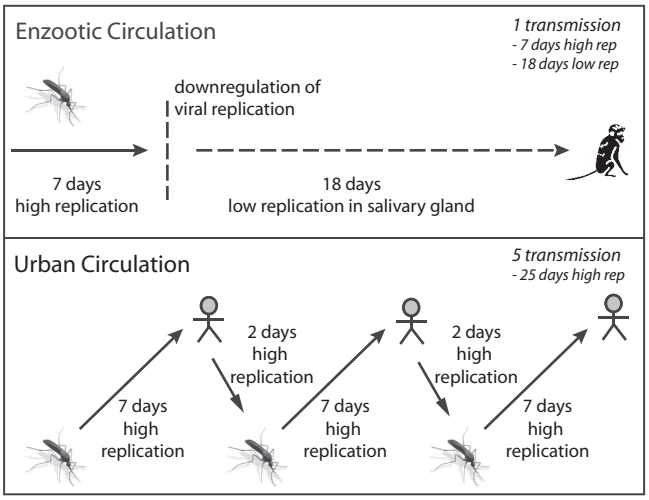

FIG. 3. Differences in the extrinsic incubation time may lead to different apparent rates of evolution. A model is depicted demonstrating how differences in vertebrate host density, availability, or herd immunity might lead to a longer extrinsic incubation time for the virus in enzootic mosquitoes, leading to less replication in the enzootic mosquitoes and resulting in the observed differences in the rate of nucleotide substitutions between enzootic and urban transmission cycles. This model assumes (i) a minimum extrinsic incubation of 1 week (blue arrows), (ii) an average intrinsic incubation (in vertebrates) of 2 days (green arrows), and (iii) downregulation of viral replication in the mosquito after 1 week (dashed arrow). (Upper panel) Sylvatic/enzootic circulation. Enzootic CHIKV is transmitted among sparsely populated or high herd immunity primates, so an infected mosquito is likely to downregulate viral replication before finding a subsequent blood-meal from a susceptible host. Less replication in a given time period results in slower evolution. (Lower panel) Urban circulation. The higher density of both competent vectors and susceptible humans that occurs in areas of urban CHIKV transmission results in more frequent transmission, meaning that an infected mosquito is likely to transmit before it downregulates viral replication. This could result in more viral RNA replication in serially infected hosts in a given time period, leading to faster evolution.

deleterious mutations and reducing the effectiveness of positive selection. Clearly, experimental infections are needed to test this hypothesis. Theoretically, immune selection could also be involved in rates of positive selection, but alphaviruses generally induce life-long immunity (55); CHIKV reinfection is therefore probably not common.

In conclusion, different patterns of CHIKV evolution in the enzootic versus epidemic transmission cycles probably reflect important differences in their ecology. Understanding these differences may be critical in preventing and controlling future CHIK outbreaks.

ACKNOWLEDGMENTS

We thank Robert Tesh and Hilda Guzman for providing CHIKV isolates. We also thank Tommy Lam for useful comments and assistance on data analysis.

S.M.V. was supported by the NIH training grant T32 AI007536. K.A.T. was supported by a James W. McLaughlin predoctoral fellowship award. P.D.M. and A.C.B. were supported by the Pacific Southwest Regional Center for Excellence U54 AI065359. SH was supported by NIH AI R21 AI073389. This study was funded by a grant to S.C.W. through the Western Regional Center of Excellence for Biodefense and Emerging Infectious Diseases (U54 AI057156) and by NIH grants AI082202 and AI069145.

The funding sources had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. **AbuBakar, S., I. C. Sam, P. F. Wong, N. MatRahim, P. S. Hooi, and N. Roslan.** 2007. Reemergence of endemic Chikungunya, Malaysia. Emerg. Infect. Dis. **13:**147–149.
2. **Adebajo, A. O.** 1996. Rheumatic manifestations of tropical diseases. Curr. Opin. Rheumatol. **8:**85–89.
3. **Adesina, O. A., and H. A. Odelola.** 1991. Ecological distribution of Chikungunya haemagglutination inhibition antibodies in human and domestic animals in Nigeria. Trop. Med. Int. Health **43:**271–275.
4. **Anonymous.** 2007. Update: chikungunya fever diagnosed among international travelers–United States, 2006. MMWR Morb. Mortal. Wkly. Rep. **56:**276–277.
5. **Arankalle, V. A., S. Shrivastava, S. Cherian, R. S. Gunjikar, A. M. Walimbe, S. M. Jadhav, A. B. Sudeep, and A. C. Mishra.** 2007. Genetic divergence of Chikungunya viruses in India (1963–2006) with special reference to the 2005–2006 explosive epidemic. J. Gen. Virol. **88:**1967–1976.
6. **Arrigo, N. C., A. P. Adams, and S. C. Weaver.** 2010. Evolutionary patterns of Eastern equine encephalitis virus in North versus South America suggest ecological differences and taxonomic revision. J. Virol. **84:**1014–1025.
7. **Bilsel, P. A., R. B. Tesh, and S. T. Nichol.** 1988. RNA genome stability of Toscana virus during serial transovarial transmission in the sandfly *Phlebotomus perniciosus*. Virus Res. **11:**87–94.
8. **Boni, M. F., D. Posada, and M. W. Feldman.** 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics **176:** 1035–1047.
9. **Brighton, S. W., O. W. Prozesky, and A. L. de la Harpe.** 1983. Chikungunya virus infection: a retrospective study of 107 cases. S. Afr. Med. J. **63:**313–315.
10. **Carey, D. E.** 1971. Chikungunya and dengue: a case of mistaken identity? J. Hist. Med. Allied Sci. **26:**243–262.
11. **Charrel, R. N., X. de Lamballerie, and D. Raoult.** 2007. Chikungunya outbreaks: the globalization of vectorborne diseases. N. Engl. J. Med. **356:**769–771.
12. **Cherian, S. S., A. M. Walimbe, S. M. Jadhav, S. S. Gandhe, S. L. Hundekar, A. C. Mishra, and V. A. Arankalle.** 2009. Evolutionary rates and timescale comparison of Chikungunya viruses inferred from the whole genome/E1 gene with special reference to the 2005–07 outbreak in the Indian subcontinent. Infect. Genet. Evol. **9:**16–23.
13. **Coffey, L. L., N. Vasilakis, A. C. Brault, A. M. Powers, F. Tripet, and S. C. Weaver.** 2008. Arbovirus evolution in vivo is constrained by host alternation. Proc. Natl. Acad. Sci. U. S. A. **105:**6970–6975.
14. **Dash, P. K., M. M. Parida, S. R. Santhosh, S. K. Verma, N. K. Tripathi, S. Ambuj, P. Saxena, N. Gupta, M. Chaudhary, J. P. Babu, V. Lakshmi, N. Mamidi, M. V. Subhalaxmi, P. V. Lakshmana Rao, and K. Sekhar.** 2007. East Central South African genotype as the causative agent in reemergence of Chikungunya outbreak in India. Vector Borne Zoonotic Dis. **7:**519–527.
15. **Deller, J. J., Jr., and P. K. Russell.** 1967. An analysis of fevers of unknown origin in American soldiers in Vietnam. Ann. Intern. Med. **66:**1129–1143.
16. **Diallo, M., J. Thonnon, M. Traore-Lamizana, and D. Fontenille.** 1999. Vectors of chikungunya virus in Senegal: current data and transmission cycles. Am. J. Trop. Med. Hyg. **60:**281–286.
17. **Drummond, A., O. G. Pybus, and A. Rambaut.** 2003. Inference of viral evolutionary rates from molecular sequences. Adv. Parasitol. **54:**331–358.
18. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.
19. **Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus.** 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. **22:**1185–1192.
20. **Edgar, R. C.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. **5:**113.
21. **Gubler, D. J.** 1988. Dengue, p. 223–260. *In* T. P. Monath (ed.), The arboviruses: epidemiology and ecology, vol. II. CRC Press, Inc., Boca Raton, FL.
22. **Hardy, J. L., E. J. Houk, L. D. Kramer, and W. C. Reeves.** 1983. Intrinsic factors affecting vector competence of mosquitoes for arboviruses. Annu. Rev. Entomol. **28:**229–262.
23. **Holmes, E. C.** 2004. The phylogeography of human viruses. Mol. Ecol. **13:**745–756.
24. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
25. **Jupp, P. G., and B. M. McIntosh.** 1990. *Aedes furcifer* and other mosquitoes as vectors of Chikungunya virus at Mica, northeastern Transvaal, South Africa. J. Am. Mosq. Control Assoc. **6:**415–420.
26. **Jupp, P. G., and B. M. McIntosh.** 1988. Chikungunya virus disease, p. 137–157. *In* T. P. Monath (ed.), The arbovirus: epidemiology and ecology, vol. II. CRC Press, Inc., Boca Raton, FL.
27. **Kariuki Njenga, M., L. Nderitu, J. P. Ledermann, A. Ndirangu, C. H. Logue, C. H. Kelly, R. Sang, K. Sergon, R. Breiman, and A. M. Powers.** 2008. Tracking epidemic Chikungunya virus into the Indian Ocean from East Africa. J. Gen. Virol. **89:**2754–2760.
28. **Khan, A. H., K. Morita, C. Parquet Md Mdel, F. Hasebe, E. G. Mathenge,**

and A. Igarashi. 2002. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. J. Gen. Virol. **83:**3075–3084.
29. **Kumarasamy, V., S. Prathapa, H. Zuridah, Y. K. Chem, I. Norizah, and K. B. Chua.** 2006. Re-emergence of Chikungunya virus in Malaysia. Med. J. Malaysia. **61:**221–225.
30. **Lahariya, C., and S. K. Pradhan.** 2006. Emergence of chikungunya virus in Indian subcontinent after 32 years: a review. J. Vector Borne Dis. **43:**151–160.
31. **Ligon, B. L.** 2006. Reemergence of an unusual disease: the chikungunya epidemic. Semin. Pediatr. Infect. Dis. **17:**99–104.
32. **Lumsden, W. H.** 1955. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53. II. General description and epidemiology. Trans R. Soc. Trop. Med. Hyg. **49:**33–57.
33. **Martin, D., and E. Rybicki.** 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics **16:**562–563.
34. **Martin, D. P., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21:**260–262.
35. **McGill, P. E.** 1995. Viral infections: alpha-viral arthropathy. Baillieres Clin. Rheumatol. **9:**145–150.
36. **Moore, D. L., S. Reddy, F. M. Akinkugbe, V. H. Lee, T. S. David-West, O. R. Causey, and D. E. Carey.** 1974. An epidemic of chikungunya fever at Ibadan, Nigeria, 1969. Ann. Trop. Med. Parasitol. **68:**59–68.
37. **Myers, R. M., and D. E. Carey.** 1967. Concurrent isolation from patient of two arboviruses, chikungunya and dengue type 2. Science **157:**1307–1308.
38. **Naresh Kumar, C. V. M., A. M. Anthony Johnson, and R. S. G. DV.** 2007. Molecular characterization of chikungunya virus from Andhra Pradesh, India, and phylogenetic relationship with Central African isolates. Indian J. Med. Res. **126:**534–540.
39. **Padidam, M., S. Sawyer, and C. M. Fauquet.** 1999. Possible emergence of new geminiviruses by frequent recombination. Virology **265:**218–225.
40. **Parola, P., X. de Lamballerie, J. Jourdan, C. Rovery, V. Vaillant, P. Minodier, P. Brouqui, A. Flahault, D. Raoult, and R. N. Charrel.** 2006. Novel chikungunya virus variant in travelers returning from Indian Ocean islands. Emerg. Infect. Dis. **12:**1493–1499.
41. **Pastorino, B., J. J. Muyembe-Tamfum, M. Bessaud, F. Tock, H. Tolou, J. P. Durand, and C. N. Peyrefitte.** 2004. Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain. J. Med. Virol. **74:**277–282.
42. **Peyrefitte, C. N., M. Bessaud, B. A. Pastorino, P. Gravier, S. Plumet, O. L. Merle, I. Moltini, E. Coppin, F. Tock, W. Daries, L. Ollivier, F. Pages, R. Martin, F. Boniface, H. J. Tolou, and M. Grandadam.** 2008. Circulation of Chikungunya virus in Gabon, 2006–2007. J. Med. Virol. **80:**430–433.
42a. **Pond, S. L., S. D. Frost, and S. V. Muse.** 2005. HyPhy: hypothesis testing using phylogenics. Bioinformatics **21:**676–679.
43. **Posada, D., and K. A. Crandall.** 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U. S. A. **98:**13757–13762.
44. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
45. **Powers, A. M., A. C. Brault, R. B. Tesh, and S. C. Weaver.** 2000. Re-emergence of Chikungunya and O'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. J. Gen. Virol. **81:**471–479.
46. **Powers, A. M., and C. H. Logue.** 2007. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. J. Gen. Virol. **88:**2363–2377.
47. **Rezza, G., L. Nicoletti, R. Angelini, R. Romi, A. C. Finarelli, M. Panning, P. Cordioli, C. Fortuna, S. Boros, F. Magurano, G. Silvi, P. Angelini, M. Dottori, M. G. Ciufolini, G. C. Majori, and A. Cassone.** 2007. Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet **370:**1840–1846.
48. **Robinson, M. C.** 1955. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53. Trans. R. Soc. Trop. Med. Hyg. **49:**28.
49. **Ross, R. W.** 1956. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. J. Hyg. **54:**177–191.
50. **Sanchez-Vargas, I., E. A. Travanty, K. M. Keene, A. W. Franz, B. J. Beaty, C. D. Blair, and K. E. Olson.** 2004. RNA interference, arthropod-borne viruses, and mosquitoes. Virus Res. **102:**65–74.
51. **Santhosh, S. R., P. K. Dash, M. M. Parida, M. Khan, M. Tiwari, and P. V. Lakshmana Rao.** 2008. Comparative full genome analysis revealed E1: A226V shift in 2007 Indian Chikungunya virus isolates. Virus Res. **135:**36–41.
52. **Schuffenecker, I., I. Iteman, A. Michault, S. Murri, L. Frangeul, M. C. Vaney, R. Lavenir, N. Pardigon, J. M. Reynes, F. Pettinelli, L. Biscornet, L. Diancourt, S. Michel, S. Duquerroy, G. Guigon, M. P. Frenkiel, A. C. Brehin, N. Cubito, P. Despres, F. Kunst, F. A. Rey, H. Zeller, and S. Brisse.** 2006. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. **3:**e263.
53. **Service, M. W.** 1965. The ecology of the tree-hole breeding mosquitoes in the Northern Guinea savanna of Nigeria. J. Appl. Ecol. **2:**1–16.
54. **Shapiro, B., A. Rambaut, and A. J. Drummond.** 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol. Biol. Evol. **23:**7–9.
55. **Smith, D. W., J. S. Mackenzie, and S. C. Weaver.** 2009. Alphaviruses, p. 1241–1274. *In* D. D. Richman, R. J. Whitley, and F. G. Hayden (ed.), Clinical virology. ASM Press, Washington, DC.
56. **Smith, G. J., D. Vijaykrishna, J. Bahl, S. J. Lycett, M. Worobey, O. G. Pybus, S. K. Ma, C. L. Cheung, J. Raghwani, S. Bhatt, J. S. Peiris, Y. Guan, and A. Rambaut.** 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature **459:**1122–1125.
57. **Smith, J. M.** 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. **34:**126–129.
58. **Swofford, D. L.** 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.
59. **Tsetsarkin, K. A., D. L. Vanlandingham, C. E. McGee, and S. Higgs.** 2007. A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog. **3:**e201.
60. **Vasilakis, N., E. C. Holmes, E. B. Fokam, O. Faye, M. Diallo, A. A. Sall, and S. C. Weaver.** 2007. Evolutionary processes among sylvatic dengue type 2 viruses. J. Virol. **81:**9591–9595.
61. **Vazeille, M., S. Moutailler, D. Coudrier, C. Rousseaux, H. Khun, M. Huerre, J. Thiria, J. S. Dehecq, D. Fontenille, I. Schuffenecker, P. Despres, and A. B. Failloux.** 2007. Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus. PLoS One **2:**e1168.
62. **Wang, E., A. C. Brault, A. M. Powers, W. Kang, and S. C. Weaver.** 2003. Glycosaminoglycan binding properties of natural Venezuelan equine encephalitis virus isolates. J. Virol. **77:**1204–1210.
63. **Yang, Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:**1586–1591.